PART3 分散と標準偏差

要点 ① 偏差と分散，標準偏差

ここでは，データの値すべてに関してその散らばり度合いをみる方法の 1 つとして，平均値との差を考える。平均値との差が大きいほど，その値は平均値から離れている。

変数 x についてのデータの n 個の値を x_1，x_2，……，x_n とし，その平均値を $\overline{x}$ とする。各値から平均値を引いた差 $x_1-\overline{x}$，$x_2-\overline{x}$，……，$x_n-\overline{x}$ を，それぞれ平均値からの「　　　」という。値が平均値より大きければ，偏差は　　であり，平均値より小さければ，偏差は　　である。

偏差の平均値を求めると

$$\frac{1}{n}\{(x_1-\overline{x})+(x_2-\overline{x})+\cdots\cdots+(x_n-\overline{x})\}=\frac{1}{n}\{(x_1+x_2+\cdots\cdots+x_n)-\overline{x}\times n\}$$

$$=\overline{\quad}-\overline{x}=$$

となるから，偏差の平均値を用いて散らばりの度合いを見ることはできない。

ここで，偏差を「　　　」した値を考えると，その値は常に　　　　となり，データの値が平均値から離れるほどその値は　　　　　なる。よって，偏差の 2 乗の平均値の大小はデータの散らばりの度合いの大小を表す 1 つの指標となると考えられる。この偏差の 2 乗の平均値を「　　　」といい，「　　」で表す。

変量 x の単位が例えば「点」であるとき，変量 x の平均値，および偏差の単位は「点」であるが，偏差の 2 乗の単位は「点^2」であるから，その平均値である分散の単位も「点^2」である。そこで，「分散の　　　　　　　　　」をとることで，変量 x の単位に合わせることができる。この「分散の正の平方根」を「　　　　　　」といい，「　　」で表す。

分散と標準偏差

分散　$s^2=\dfrac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}$

標準偏差　$s=\sqrt{s^2}=\sqrt{\dfrac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}}$

例　5 人の生徒の小テストの得点 x が右の表で与えられているとき，分散と標準偏差を求める。

x	6	10	8	7	4
$x-\overline{x}$					
$(x-\overline{x})^2$					

平均値は $\overline{x}=\frac{1}{5}(6+10+8+7+4)=\frac{35}{5}=7$（点）

分散は $s^2=\frac{1}{5}\{\qquad\qquad\qquad\qquad\qquad\}=\frac{1}{5}(1+9+1+0+9)=\frac{20}{5}=4$

標準偏差は $s=\sqrt{s^2}=\sqrt{4}=2$（点）

データの値が平均値の周りに集中しているほど，分散や標準偏差は小さくなる。よって，複数のデータの標準偏差を比較することによって，散らばりの度合いの比較をすることができる。

※データの中に，他の値と極端に離れた値がある場合，その値の偏差の 2 乗の値が極端に大きくなり，分散や標準偏差の値も大きくなる。

要点 ② 分散の式の変形

分散 s^2 は，次のように変形することができる。

$$s^2=\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}$$

$$=\frac{1}{n}[\{x_1{}^2-2x_1\overline{x}+(\overline{x})^2\}+\{\qquad\qquad\qquad\}+\cdots\cdots+\{x_n{}^2-2x_n\overline{x}+(\overline{x})^2\}]$$

$$=\frac{1}{n}\{(x_1{}^2+x_2{}^2+\cdots\cdots+x_n{}^2)-2\overline{x}(x_1+x_2+\cdots\cdots+x_n)+\quad(\overline{x})^2\}$$

$$=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots\cdots+x_n{}^2)-2\overline{x}\cdot\frac{1}{n}(x_1+x_2+\cdots\cdots+x_n)+(\overline{x})^2$$

$$=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots\cdots+x_n{}^2)-2\overline{x}\cdot\quad+(\overline{x})^2=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots\cdots+x_n{}^2)-(\overline{x})^2$$

よって，次のことがいえる。

変量 x の分散

$s^2=(x^2$ の平均値$)-(x$ の平均値$)^2=\overline{x^2}-(\overline{x})^2$

例　5 人の生徒の小テストの得点 x が右の表で与えられているとき，分散 s^2 を求める。

x	6	10	8	7	4

$$\overline{\quad}=\frac{1}{5}(\quad+\quad+\quad+\quad+\quad)=\frac{265}{5}=53$$

$$\overline{x}=\frac{1}{5}(6+10+8+7+4)=\frac{35}{5}=7$$

よって　$s^2=\quad-\quad=4$

3　次のデータは，8 人の生徒が 2 か月の間に読んだ本の冊数 x(冊)である。

1，3，6，5，8，1，6，2

(1)　このデータの平均値 $\overline{x}$ を求めよ。

(2)　このデータの分散 s^2 を求めよ。

(3)　このデータの標準偏差 s を求めよ。

Point Pickup

分散と標準偏差

変量 x のデータの n 個の値が x_1，x_2，……，x_n，その平均値を $\overline{x}$ とするとき

分散　$s^2=\dfrac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}$　……(偏差)^2 の平均値

標準偏差　$s=\sqrt{s^2}=\sqrt{\dfrac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}}$　……$\sqrt{(分散)}$

分散は

$s^2=\overline{x^2}-(\overline{x})^2$　……　(x^2 の平均値)−(x の平均値)^2

と計算することができる。

※データの中に他の値と極端に離れた値がある場合，分散や標準偏差の値も大きくなる。

第 15 講　PART3 確認問題

次の 8 個のデータ x がある。

1，2，2，3，4，6，7，7

1　このデータの平均値 $\overline{x}$ は $\boxed{\text{ア}}$ である。

2　このデータの分散 s^2 は $\boxed{\text{イ}}$ である。

3　このデータの標準偏差 s は $\sqrt{\boxed{\text{ウ}}}$ である。

PART3 分散と標準偏差

3 次のデータは，8 人の生徒が 2 か月の間に読んだ本の冊数 x(冊)である。

1，3，6，5，8，1，6，2

(1) このデータの平均値 $\overline{x}$ を求めよ。

(2) このデータの分散 s^2 を求めよ。

(3) このデータの標準偏差 s を求めよ。

[解答]

(1) $\overline{x}=\dfrac{1}{8}(1+3+6+5+8+1+6+2)=\dfrac{32}{8}=4$ (冊)

(2) $s^2=\dfrac{1}{8}\{(-3)^2+(-1)^2+2^2+1^2+4^2+(-3)^2+2^2+(-2)^2\}=\dfrac{48}{8}=6$

(別解) $\overline{x^2}=\dfrac{1}{8}(1^2+3^2+6^2+5^2+8^2+1^2+6^2+2^2)=\dfrac{176}{8}=22$

よって，$s^2=\overline{x^2}-(\overline{x})^2=22-4^2=6$

(3) $s=\sqrt{s^2}=\sqrt{6}$ (冊)

第15講 PART3 確認問題 解答

1

[正解] ア：4

[解説]

平均値 $\overline{x}$ は，[A]

$$\overline{x}=\frac{1}{8}(1+2+2+3+4+6+7+7)$$

$$=\frac{1}{8}\times 32=\mathbf{4}$$

> [A]　平均値
> $\overline{x}=\frac{1}{n}(x_1+x_2+\cdots+x_n)$

2

[正解] イ：5

[解説]

データの各値を x とすると，平均値 $\overline{x}=4$ であり，

$x-\overline{x}$ と $(x-\overline{x})^2$ の値を求めると，次のようになる。

x	1	2	2	3	4	6	7	7
$x-\overline{x}$	-3	-2	-2	-1	0	2	3	3
$(x-\overline{x})^2$	9	4	4	1	0	4	9	9

よって，分散 s^2 は，[B]

$$s^2=\frac{1}{8}(9+4+4+1+0+4+9+9)$$

$$=\frac{1}{8}\times 40=\mathbf{5}$$

> [B]　分散
> s^2
> $=\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots$
> $+(x_n-\overline{x})^2\}$
> $s^2=\overline{x^2}-(\overline{x})^2$
> $(x^2$ の平均値$)-(x$ の平均値$)^2$
> と計算することもできる。

3

[正解] ウ：5

[解説]

分散 s^2 が 5 であるから，標準偏差は，[C]

$$s=\sqrt{s^2}=\sqrt{\mathbf{5}}$$

> [C]　標準偏差 $s\cdots\sqrt{(\text{分散})}$